Electronic Journal of Applied Statistical Analysis
EJASA (2012), Electron. J. App. Stat. Anal., Vol. 5, Issue 3, 339 – 345
e-ISSN 2070-5948, DOI 10.1285/i20705948v5n3p339

PERMUTATION TEST FOR GROUP COMPARISON IN PLS-PATH MODELING

Anna Crisci*, Antonello D'Ambra

Department of Economics, Second University of Naples, Italy

Received 16 July 2012; Accepted 07 October 2012
Available online 16 November 2012

Abstract*: In the present paper we consider a group comparison in PLS-Path Modeling. The aim of this methodology is to verify whether significant differences between the two groups in terms of path coefficient exist. The method used to test such differences is the Permutation test, that is a test of randomization that provides a non- parameter option. Using real data, we show how the permutation procedure can be used to understand whether the two groups differ.*

Keywords*: Structural equation model, group-comparison, permutation procedure, randomization test.*

1. Introduction

The advanced procedures to compare multiple-groups were implemented in SEM-Covariance-Based. The Multi-groups Structural Equation Models allow you to examine models simultaneously across multiple samples. This happens through the study of the invariance [11,12,14] which proceeds sequentially through a series of steps each of which introduces additional constraints with respect to the initial model. In the absence of constraints between groups each group can be analyzed separately, while in presence of constraints between groups the data of all groups must be analyzed simultaneously. The basic requirement for a model of multiple groups is that populations are clearly defined and the samples are independent (i.e males and females). By means of multi-group models, any assumptions concerning the invariance can be examined, considering as extreme assumptions those in which:

- all parameters are not invariant (there are not constraints on parameters);
- all parameters are invariant (all parameters are constrained);

* E-mail: crisci.anna@virgilio.it

In the present paper we consider a group comparison in PLS-Path Modeling. The aim is to verify whether significant differences between the two groups[1] in terms of path coefficient exist. The method to verify such significant differences is the Permutation test, that is a test of randomization that provides a non- parameter option.
The paper is organized as follow: in Section 2, the permutation procedure is shown. In Section 3 a case study will is shown and in Section 4 a discussion is shown.

2. Group Comparison in PLS-Path Modeling

As previously alluded, the advanced procedures to compare multiple-groups were implemented in SEM-Covariance- Based. This approach, however, implies numerous questions about the propriety of the data and the size of the sample. Another less restrictive way to test structural equation models between groups is the use of Partial Least Squares (PLS).
The researchers often examine and discuss just the difference in the size of the estimates of the paths of two or more sets of the data. When we estimate the meaning of the differences of the paths of a particular model for two or more sets of the data, a *t-test* based on the standard errors is obtained by means of a re-sampling procedure like bootstrap[2]. Yet, problems may rise if the assumption of a normal population or of a similar group sample size is not met.
An alternative approach, i.e a permutation or randomization procedure [3,8] is available, in which a subset of all the possible data permutations of the data between the sample groups is constructed.

2.1 The Permutation Procedure

Randomization or permutation procedures are the favorite tests of significance for non- normal data. These techniques are considered *distribution –free tests* in that they require no parametric assumptions. Randomization tests should not be viewed as alternative to parametric statistical tests; rather, they should be considered as tests for that particular empirical form to be examined.
The procedure for a permutation test based on random assignment, as described by Edington (1987) [7] and Good (2000) [9], and subsequently illustrated by Chin and Dibbern (2010)[4], is carried out in the following way:

1. A test statistic is computed for data.
2. The data are permuted (divided or re-arranged) repeatedly in a way consistent with the random assignment procedure. With two or more samples, all observations are combined into a single large sample before being rearranged. The test statistic is computed for each of the resulting data permutations;
3. The proportion of the permutations of the data in the set of reference having the values of the test statistic ≥ (or for some statistic tests, ≤) to the value of the results obtained experimentally is the ***P-value***, that is the minimal level of significance to which it is possible to reject the null hypothesis.

---

[1] Adding just one more group results in a dramatic increase in the number of permutations, and the numbers are so dramatically high that they will forever be out of reach of computers [5]

[2] The bootstrap method:

1) Compute the difference in the parameter for two groups (i.e. the difference between path coefficients);
2) Separate the data into groups and run bootstrap re-samplings for each group. Path coefficients are calculated in each re-sampling and the standard error estimated are treated in a parametric sense via t-tests [2]

When the basis for the permutation of the data is the random attribution, the permutation test is often referred to as "***Randomization- test***". This previous definition is wide enough to include procedures called *randomization tests* that depend on both random samples and randomization. The modern concept of randomization is, however, a permutation test which is just based on randomization, where it is not important the way by which the sample has been chosen. As Edington (1987) underlines, a permutation test based on randomization "is valid for any type of sample, regardless of the way the sample is chosen".
The null and alternative hypothesis to be tested to compare the PLS parameter (path coefficients) estimations between two independent groups G1 $(m_1,m_2,\ldots,m_l)$ and G1 $(m_1,m_2,\ldots,m_k)$ are H0: path coefficients are not significantly different; H1: path coefficients are significantly different.

3. The case study

In this study, the innovative performance of the Manufacturing Enterprises in Campania has been evaluated. To evaluate the degree of innovation achieved by enterprises, the survey data UniCredit-2007 [18] concerning Campania have been used. The enterprises under study are those of the first, second and third sectors based on Pavitt classification [13]: **supplied dominated**, **scale intensive, specialized supplied**. In particular, the enterprises of different sectors have been compared by means of the group comparison. The LVs and their corresponding MVs identified in the present study are:

Table 1. Latent and Manifest Variables.

Latent Variables	Manifest Variables
Company Size	• Turnover • Number of employees;
Investment Activity	• Fixed asset costs (Plants, Machinery and equipment); • Information and Communication Technology costs;
Human Capital	Sub-division of employed with qualification: • Primary and Secondary schools; • High school- certificate/degree; • Number of employees with a degree in 2006; • Employees in R&D;
Innovative Activity	• Costs in R&D internal; • Costs in R&D external; • Other costs in technology Innovation (purchase of plants, equipment designed for the introduction of new products/processes, purchase of other technology, innovative marketing products etc.)
Degree of Innovation Diffusion	• Product innovation; • Processes innovation; • Contextual adoption of product and process innovation.

The existence of meaningful relations between the following LVs has been measured:

1) *Company Size and Innovation Activity;*
2) *Investment Activity and Innovation Activity;*
3) *Company Size and Investment Activity;*
4) *Human Capital and Degree of diffusion Innovation*

5) Innovation Activity and Degree of diffusion Innovation;

Before proceeding to the estimation of the parameters we have verified the unidimensionality of the VMs blocks by means of Dillon-Goldstein's rho [6].A blocks is unidimensionality if this index is > 0.7. The value of the index is > 0.7 for all the observed VMs blocks.
To estimate the parameters of the model, we have used the module R-package [19]. The variables have all been standardized. To calculate the inner estimates of the latent variables, we have used the path weighting scheme. The table 2 shows the bootstrap results for path coefficients[1].The bootstrapping analysis allows for the statistical testing of the hypothesis H0:w=0 (w can be any parameter estimated by PLS) against the alternative hypothesis H1:w≠0. In particular, bootstrap has been based on 100 samples. From bootstrap analysis we infer a significance of all the links (T-statistics >2) except the link between the Investment Activity and Innovative Activity (T-statistics < 2).

Table 2. Bootstrap results for path coefficients.

			Original	Mean.Boot	Std.Err	T-Test
Company Size	→	Investment Activity	0.72	0.71	0.14	5.14**
Company Size	→	Innovative Activity	0,69	0.62	0.15	4.60**
Investments Activity	→	Innovative Activity	0.09	0.11	0.18	0.50
Human Capital	→	Degree of Diffusion Innovation	0.28	0.29	0.07	4.00**
Innovative Activity	→	Degree of Diffusion Innovation	0.37	0.39	0.08	4.63**

***significant 5%*

3.1 Group Comparison: Permutation test

To compare the enterprises of the first, second and third sectors according to Pavitt Classification, two groups have been created:

- Group.1: all the enterprises of the first sector (***supplied dominated)***- 43 units- belong to this group; .
- Group. 2: the enterprises of the second ***(scale intensive***) and third ***(specialized supplied)*** sectors- 47 units- have been considered in this group.

As it is well- known, the aim of this methodology is to verify whether significant differences between the two groups in terms of path coefficient exist. The method used to test such differences is the Permutation test. The table 3 shows the obtained results. The first column shows the path coefficients for Group1, the second column the path coefficients for Group2. Likewise, the third column contains the absolute difference of path coefficient between the two groups. In contrast, the four column has the p-value of permutation test. In particular, the significantly different path coefficients (p-value of the permutation test < 0.05) are those in bold. The first significant difference may be explained by the fact that the enterprises of the second group (scale intensive/specialized supplied) are those that give more importance to the fixed investment. As concerns the second significant difference, instead, the enterprises of the traditional sectors (supplied-dominated) actually attract more labor and provide a higher weight of the human capital on the degree of diffusion innovation (Figure 1).

Table 3. Group Comparison: Permutation Test.

			G1	G2	Diff.Abs	p-value
Company Size	→	Investments Activity	**0.53**	**0.89**	**0.36**	0.04**
Company Size	→	Innovative Activity	0.73	0.72	0.01	0.98
Investment Activity	→	Innovative Activity	0.10	0.03	0.07	0.83
Human Capital	→	Degree of Diffusion Innovation	**0.34**	**-0.08**	**0.42**	0.01**
Innovative Activity	→	Degree of Diffusion Innovation	0.36	0.4	0.04	0.85

***significant 5%*

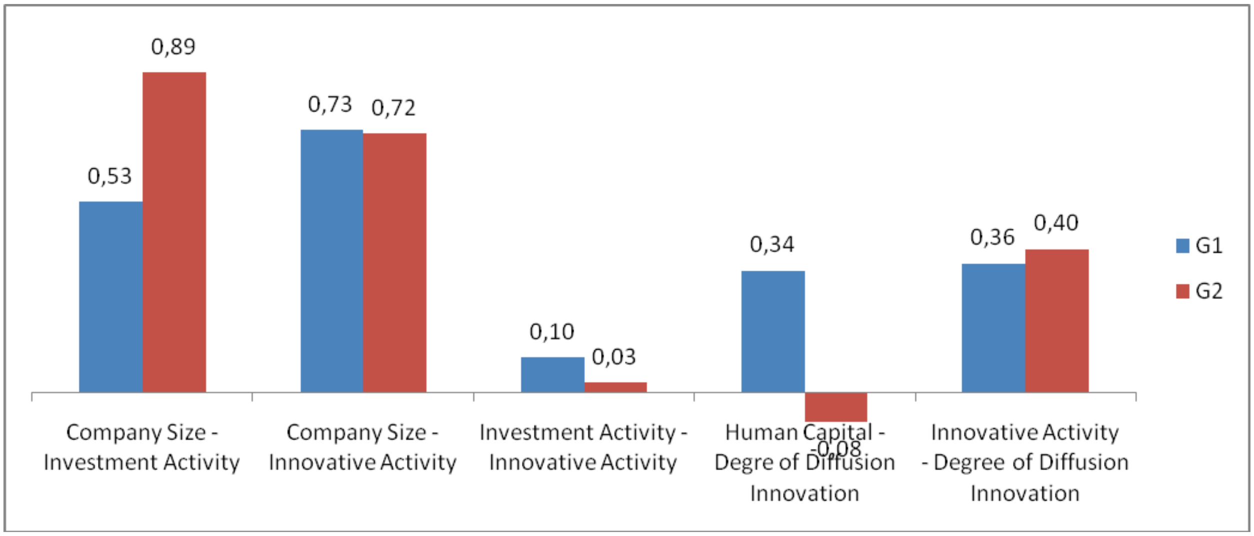

Figure 1. Group Comparison.

4. Discussion

The aim of the present work has been to show, by means a real application, the method to compare two independent groups in the context of PLS-Path Modeling. The method to verify whether significant differences in term of path coefficients exist is the Permutation test or randomization procedure. This procedures are the favorite test of significant for non-normal data. Permutation tests exist for any test statistic, regardless of whether or not its distribution is known. Permutation tests can be used for analyzing unbalanced designs and for combining dependent tests on mixtures of categorical, ordinal, and metric data [15]. Attractive as they are, permutation tests also have drawbacks. Are primarily used to provide p-value and their implementation requires a case-by-case examination, with new codes to be written for each specific problem. Permutation tests are computationally heavy; in general, complete enumeration of $n!$ permutations is infeasible, and a sampling strategy has to be adopted. This latter aspect indicates similarities between permutation tests and other re-sampling techniques such as bootstrap; contrary to bootstrap-based tests, however, permutation tests, irrespective of the sample size, are exact tests [16].
Good (2000) explains the difference between permutation tests and bootstrap tests the following way: "Permutations test hypotheses concerning distributions; bootstraps test hypotheses concerning parameters.

Recent monographs, providing theoretical results and large numbers of applications along with softwares, are [10] and [17].

Acknowledgement

The authors thank reviewers for the suggestions.

References

[1]. Chin, W.W. (1998). The partial least squares approach to structural equation modeling. *Modern Methods for Business Research*, 295-336. Marcoulides, G.A. (Ed). London: Lawrence Erlbaum Associates.

[2]. Chin,W.W. (2000). Frequently Asked Questions, Partial Least Squares PLS-Graph. Available from: http: //disc-nt.cba.uh.edu/chin/plsfaq/multigroup.htm.

[3]. Chin, W.W. (2003). A permutation procedure for multi-group comparison of PLS models. M. J. Vilares, M., Tenenhaus, M., Coelho, P., Esposito Vinzi, V., Morineau, A. *PLS and Related Methods* - Proceedings of the International Symposium PLS03. Decisia, pp. 33-43.

[4]. Chin, W.W & Dibbern, J. (2010). An introduction to a Permutation Based Procedure for Multi-Group PLS Analysis. *Handbook of Partial Least Squares*. Springer, part 1, 171-193.

[5]. Cousineau D. (2011). Randomization test of mean is computationally inaccessible when the number of groups exceeds two. *Tutorial in Quantitative Methods for Psychology*, vol. 7(1), pp. 15-18.

[6]. Dillon,W.R. & Goldestein, M. (1984). *Multivariate Analysis: Method and Applications.* Wiley, New York.

[7]. Edgingthon, E. (1987). *Randomization test*, Marcel Dekker Inc.

[8]. Esposito Vinzi, V., Chin, W.W., Henseler,J., and Wang. H (2010). *Handbook of Partial Least Squares, Concepts, Methods and Application*. Springer, New York.

[9]. Good, P.(2000). *Permutation test: a practical guide to resampling for testing hypothesis*. Springer, New York.

[10]. Good, P. I. (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses*, 3rd edition. New York: Springer.

[11]. Horn, JL., McArdle, JJ.(1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18:117-144.

[12]. Hui, C. Harry and Harry, C. Triandis (1985). Measurement in Cross-Cultural Psychology: A Review and Comparison of Strategies. *Journal of Cross-Cultural Psycology*, 16 (June), 131-152.

[13]. Malerba, F. (2000). *Economia dell'innovazione*, Carocci Editore.

[14]. Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance, *Psychometrika*, 56:525-543.

[15]. Pesarin, F. (2001). *Multivariate Permutation Tests*. New York: John Wiley & Sons.

[16]. Pesarin F., Salmaso L. (2006). Permutation test for univariate and multivariate ordered categorical data. *Austrian Journal of Statistics*, vol. 35, 2&3,pp. 315-324.

[17]. Pesarin, F. and Salmaso, L. (2010). *Permutation Tests for Complex Data: Theory, Applications and Software*. New York: John Wiley & Sons.

[18]. Rapporto UniCredit - Corporate Banking (2007). Decima indagine sulle Imprese Manifatturiere.
[19]. Sanchez, G. (2009). Partial Least Squares (PLS) data Analysis methods. Package.